

University of Pennsylvania Carey Law School

Penn Law: Legal Scholarship Repository

Faculty Scholarship at Penn Law

1996

An Analysis of Fee Shifting Based on the Margin of Victory: On Frivolous Suits, Meritorious Suits, and the Role of Rule 11

Howard F. Chang

University of Pennsylvania Carey Law School

Lucian A. Bebchuk

Follow this and additional works at: https://scholarship.law.upenn.edu/faculty_scholarship



Part of the [Civil Law Commons](#), [Civil Procedure Commons](#), [Economic Policy Commons](#), [Economic Theory Commons](#), [Law and Economics Commons](#), [Legal Theory Commons](#), and the [Litigation Commons](#)

Repository Citation

Chang, Howard F. and Bebchuk, Lucian A., "An Analysis of Fee Shifting Based on the Margin of Victory: On Frivolous Suits, Meritorious Suits, and the Role of Rule 11" (1996). *Faculty Scholarship at Penn Law*. 969. https://scholarship.law.upenn.edu/faculty_scholarship/969

This Article is brought to you for free and open access by Penn Law: Legal Scholarship Repository. It has been accepted for inclusion in Faculty Scholarship at Penn Law by an authorized administrator of Penn Law: Legal Scholarship Repository. For more information, please contact PennlawIR@law.upenn.edu.

AN ANALYSIS OF FEE SHIFTING BASED ON THE MARGIN OF VICTORY: ON FRIVOLOUS SUITS, MERITORIOUS SUITS, AND THE ROLE OF RULE 11

LUCIAN ARYE BEBCHUK and HOWARD F. CHANG*

ABSTRACT

When plaintiffs cannot predict the outcome of litigation with certainty, neither the American rule (each litigant bears its own litigation expenses) nor the British rule (the losing litigant pays the attorneys' fees of the winning litigant) would induce optimal decisions to bring suit. Plaintiffs may bring frivolous suits when litigation costs are small relative to the amount at stake; plaintiffs may not bring meritorious suits when litigation costs are large relative to this amount. More general fee-shifting rules are based not only on the identity of the winning party but also on how strong the court perceives the case to be at the end of the trial (the "margin of victory"). We analyze when such a rule can induce plaintiffs to sue if and only if they believe their cases are sufficiently strong. We explore the implications of our analysis for the use of Federal Rule of Civil Procedure 11.

I. INTRODUCTION

UNDER fee-shifting rules a court can require the losing litigant to pay the attorneys' fees of the winning litigant. Although there exists a substantial literature on the economic analysis of such fee-shifting rules, the assumption throughout has been that fee shifting will depend only on the identity of the winning party. This article contributes to that literature by

* Professor of Law, Economics, and Finance, Harvard Law School, and Associate Professor of Law, University of Southern California Law School, respectively. We wish to thank Scott Altman, Dick Craswell, Franklin Fisher, Ron Garet, Andrew Guzman, Oliver Hart, Avery Katz, Greg Keating, George Lefcoe, Ed McCaffery, Mitch Polinsky, Eric Rasmusen, Judith Resnik, Steven Salop, Steven Shavell, Warren Schwartz, Matt Spitzer, and an anonymous referee, as well as seminar participants at the University of Southern California, at Georgetown University, and at the 1995 meeting of the American Law and Economics Association, for helpful comments and the John M. Olin Foundation for financial support. Lucian Bebchuk's work has also been supported by the National Science Foundation. Howard Chang's work has also been supported by the Conrad N. Hilton Endowed Fund for the Improvement of the Administration of Justice and the James H. Zumberge Faculty Research and Innovation Fund.

considering the effect of fee-shifting rules that are based not only on which party won the case but also on how strong the court perceives the plaintiff's case to be at the end of the trial—that is, the “margin of victory.” In particular, we analyze how expanding the set of instruments available to courts can provide better incentives for a plaintiff to bring suit. This analysis reveals how and when one can design such a rule to induce plaintiffs to sue only if they believe their cases are sufficiently strong.¹

This analysis of fee-shifting rules based on the margin of victory is not only of theoretical interest but also of practical significance: as we will discuss in Section VII, courts have interpreted Federal Rule of Civil Procedure 11 as an example of such a rule. In order to deter parties from filing frivolous papers in court, Rule 11 authorizes courts to impose sanctions on those who file such papers. Typically, a court uses Rule 11 against the plaintiff when it determines that the plaintiff's claims are so lacking in merit that they are frivolous; the sanction that courts have imposed most frequently is an order that the plaintiff pay the defendant's attorney's fees. Our analysis will produce conclusions that shed light on the role that Rule 11 can play in improving the plaintiff's incentives to bring suit.²

Suppose that society seeks to induce plaintiffs to sue if and only if they believe they are entitled to prevail at trial. (As discussed in Section VI, the analysis of this article is more general, that is, also allows for other objectives.) This objective implies two goals: that a plaintiff bring a “meritorious” suit—which we define as a suit that deserves to win on the merits, as the plaintiff views the case—and that a plaintiff not bring a “frivolous” suit—which we define as a suit that does not deserve to win on the merits, as the plaintiff views the case. How can fee-shifting rules best serve these goals?³

¹ For our purposes, we will treat the plaintiff and its attorneys as one actor. That is, we do not introduce principal-agent problems between each party and its attorneys.

² Although we will describe our analysis in terms of a plaintiff's incentives to bring suit, the model also generalizes to apply to any party's incentives to file any paper, with fee shifting based on the costs of litigating the merits of that paper and on the margin of victory on that particular issue. Thus, our analysis extends to the use of Rule 11 sanctions in these other contexts, including the use of sanctions against the defendant. The Federal Rules of Civil Procedure also provide other examples of fee shifting based on the margin of victory. For example, Rule 37(a)(4) requires a court to award attorney's fees to the winning party on a motion for an order compelling disclosure or discovery, unless a losing party “was substantially justified” in making the motion (in the case of a losing movant) or in opposing the disclosure or discovery (in the case of a winning movant).

³ A. Mitchell Polinsky & Daniel L. Rubinfeld, *Sanctioning Frivolous Suits: An Economic Analysis*, 82 *Geo. L. J.* 397 (1993), and A. Mitchell Polinsky & Daniel L. Rubinfeld, *Optimal Awards and Penalties When the Probability of Prevailing Varies among Plaintiffs*, 27 *RAND*

One important consideration is the fact that the outcome of a trial is unlikely to be certain to the plaintiff when it decides whether to sue. One source of uncertainty is that the court might lack information available to the plaintiff or can err in its judgment given the information that it can observe. Furthermore, the plaintiff might lack information that a trial would later reveal to the court. As a result, the plaintiff might attach some probability to winning even if it views its case as weak, and it might attach some probability to losing even if it views its case as strong.⁴

Consider first the standard American rule, under which each litigant bears its own expenses. This rule does not induce optimal litigation decisions. First, plaintiffs will not bring all meritorious suits. Even if the plaintiff can count on the court to decide the case as the plaintiff predicts, the plaintiff will not sue if its litigation costs exceed the value of the relief that it expects the court to award. Second, if the plaintiff believes that the court's decision might differ from what the plaintiff expects, either because the court might err or because the plaintiff might err, then as we will show, the plaintiff will bring some frivolous suits. If the litigation costs are small enough, the plaintiff will find it worthwhile to gamble—either because the court might err in the plaintiff's favor or because the case might prove to be better than it first appears.

Consider next the British rule, under which the loser pays the expenses of the winner. If the plaintiff could predict the trial outcome without error, then the plaintiff would not bring a frivolous suit and would never be discouraged by litigation expenses from bringing a meritorious suit, which would guarantee the plaintiff reimbursement of its litigation expenses. If the plaintiff cannot predict the trial outcome with certainty, however, then as we will show, the plaintiff will not always make litigation decisions consistent with the goals suggested above. First, if litigation costs are small enough, then the plaintiff will bring some frivolous suits because it might win at trial anyway. Second, if litigation costs are sufficiently large, then the prospect of losing (and bearing the expenses of both litigants) will deter the plaintiff from bringing some meritorious suits because it still might lose at trial.

J. Econ. 269 (1996), examine how awards and sanctions can discourage frivolous suits and encourage meritorious suits. Polinsky and Rubinfeld do not, however, consider the subject that is the main focus of this article: how fee shifting based on the margin of victory can advance these goals. Whereas we focus on the criteria courts might use in deciding whether or not to shift fees, Polinsky and Rubinfeld focus on the magnitudes of awards and sanctions as policy instruments.

⁴ A. Mitchell Polinsky & Steven Shavell, *Legal Error, Litigation, and the Incentive to Obey the Law*, 5 J. L. Econ. & Org. 99 (1989), stressed the important effect of legal uncertainty caused by judicial error on the incentive to sue. Polinsky and Shavell do not study, however, how fee shifting based on the margin of victory can address the problem of judicial error (and other reasons for the unpredictability of judgment).

This article shows how to design fee-shifting rules so as to induce better litigation decisions. The fee shifting should depend not only on which party prevails but also on the margin by which they prevailed. That is, the rule should take into account not only who won but also the degree to which they won easily. This information is useful because a plaintiff who loses by a large margin is less likely to have believed *ex ante* that the case was meritorious than a plaintiff who loses by a small margin. Similarly, a plaintiff who wins by a large margin is less likely to have believed *ex ante* that the case was frivolous than a plaintiff who wins by a small margin.

Given the distribution of the errors in the plaintiff's prediction of the trial outcome, it is possible to design a fee-shifting rule that would induce better litigation decisions. Because the optimal rule depends on this distribution, a court can implement such a rule as long as it has some sense of how these errors are distributed. This article examines the structure of such rules and how they affect litigation decisions, then applies our conclusions to Rule 11 as it exists today.

Section II of this article presents the formal model of litigation that we shall use to analyze the plaintiff's incentives to bring suit. Section III examines those incentives under the classic fee-shifting rules, which allocate litigation costs according to the identity of the losing party. In Section IV, we allow the fee-shifting rule to depend on the margin of victory as well, and we show how one can thereby provide the plaintiff with better incentives to bring suit. Section V shows that if the court can shift the fees of either litigant in each case, then there exists a whole family of such fee-shifting rules that can present plaintiffs with these improved incentives. Section VI explores some extensions of our model. Section VII considers the implications of our analysis for the application of Rule 11. Section VIII summarizes our conclusions.

II. FRAMEWORK OF ANALYSIS

Consider a potential plaintiff who is deciding whether to bring suit against a potential defendant. Let x be the parameter relevant for determining the merits of the plaintiff's suit (or an index that summarizes all the relevant parameters). Without loss of generality, we can define x such that the plaintiff's case is stronger, and the defendant's case is correspondingly weaker, as x increases. In particular, let \bar{x} represent the threshold value for a victory for the plaintiff: the court will decide in favor of the plaintiff if and only if the court finds that x exceeds \bar{x} .

This framework is general enough to describe the situation under a

wide variety of rules setting forth the standard of liability. For example, the defendant's liability in a particular tort suit might depend on whether the plaintiff was contributorily negligent, in which case x would represent the amount by which the plaintiff's level of care exceeded the standard for negligence, and \bar{x} would equal zero. Or the case might turn on whether the defendant was negligent, in which case x would represent the amount by which the defendant's actual level of care fell short of the standard of due care (that is, the extent of the defendant's underinvestment in safety precautions), and \bar{x} would equal zero. Or the case might turn on an issue of causation, and x might be a feature that determines whether the court will find proximate cause.

Let x_c denote the value of x observed by the court and thus the value of x on which the court bases its judgment. An important feature of our model is that the court may not be able to identify x accurately. Specifically, except in Section IIIA, we assume that the court errs by a random amount ϵ , so that $x_c = x + \epsilon$. The random variable ϵ is distributed according to a cumulative distribution function $F(\epsilon)$. For any particular realization of ϵ , ϵ' , the corresponding $F(\epsilon')$ will equal the probability *ex ante* that $\epsilon \leq \epsilon'$. Let ϵ take on positive and negative values, each with some positive probability, so that $0 < F(0) < 1$. For simplicity, assume that ϵ only takes on values within the interval $(-e_1, e_2)$, where both e_1 and e_2 are positive, and let $F(\epsilon)$ be continuous and strictly increasing over this interval. Let F^{-1} denote the inverse function of $F(\epsilon)$, defined over the domain $[0, 1]$ such that $F^{-1}(0) = -e_1$ and $F^{-1}(1) = e_2$.

Let θ_p denote the probability that $\epsilon > 0$, that is, the likelihood that the court will err in favor of the plaintiff, so that $\theta_p = 1 - F(0)$. Accordingly, θ_p is the likelihood that a defendant who just barely deserves to win on the merits (that is, a defendant in the marginal case in which $x = \bar{x}$) will nevertheless lose because the court errs in favor of the plaintiff. Let θ_d denote the probability that $\epsilon \leq 0$, that is, the likelihood that the court will not err in favor of the plaintiff, so that $\theta_d = F(0)$. Accordingly, θ_d is the likelihood that the defendant in the marginal case (in which $x = \bar{x}$) will win.

The case in which $\theta_d = \theta_p = 1/2$ will prove to be a useful example. This case is symmetric insofar as the court is as likely to decide the marginal case in favor of the defendant as it is in favor of the plaintiff: the court is as likely to overestimate the strength of the plaintiff's case as it is to underestimate it. Our framework is not limited to the symmetric case, however, and covers also those cases in which the court systematically errs in favor of one side.

The plaintiff expects the court to require the defendant to pay the

plaintiff some amount if the court finds the defendant liable. Let D denote the expected award conditional on a finding of liability, where $D > 0$.⁵ (For example, in a tort case, D is the expected amount of damages that the court would award to the plaintiff if the defendant is found liable.) We assume that D , the expected amount at stake, is independent of x_c . If D is instead a function of x , then the $D(x_c)$ function is another policy instrument, which we consider as an extension of our model in Section VIC. Let C_p and C_d be the litigation costs of the plaintiff and of the defendant, respectively, where $C_p > 0$ and $C_d > 0$.

Let x_p be the plaintiff's observation of x when deciding whether to bring suit. At this stage, we assume that the plaintiff knows the true value of x ; that is, $x_p = x$. Therefore, $x_c = x_p + \epsilon$. Section VIA explains how the analysis can be extended easily to the case in which the plaintiff is imperfectly informed about x . In particular, the court may have information available at trial that is unavailable to the plaintiff when deciding whether to bring suit.

The plaintiff is risk neutral. For simplicity, we assume that the plaintiff sues if and only if the expected value of bringing the suit and going to trial is positive.⁶ The court observes x_c before reaching judgment and also knows D , $F(\epsilon)$, C_p , and C_d when it applies the fee-shifting rule. When the plaintiff decides whether to sue, it knows x_p , D , $F(\epsilon)$, C_p , C_d , and the fee-shifting rule.

Suppose that the social objective is to induce the plaintiff to sue if and only if $x_p > x^*$, where x^* is some threshold. This general statement of the objective is reasonable: we cannot have the plaintiff act on anything but its impression of its case, and presumably if we want to deter suits for a certain value of x_p , we also want to deter suits for any lower values. At this stage, we assume that $x^* = \bar{x}$. That is, the plaintiff should sue if and only if the plaintiff believes that the defendant is liable. Thus, to get optimal decisions to bring suit, the expected value of going to trial must

⁵ This model also generalizes to include suits in which the plaintiff seeks injunctive relief, if we let D include the expected value of this relief to the plaintiff.

⁶ If parties can settle before going to trial, then plaintiffs might find it worthwhile to bring suits with negative expected value, solely to extract a settlement offer. See Lucian A. Bebchuk, *Suing Solely to Extract a Settlement Offer*, 17 J. Legal Stud. 437 (1988); Lucian A. Bebchuk, *A New Theory concerning the Credibility and Success of Threats to Sue*, 25 J. Legal Stud. 1 (1996); David Rosenberg & Steven Shavell, *A Model in Which Suits Are Brought for Their Nuisance Value*, 5 Int'l Rev. L. & Econ. 3 (1985). Plaintiffs will bring every suit with positive expected value, whether or not settlements are possible. Thus, the possibility of settlements raises the incentives for plaintiffs to sue, compared with a model that excludes settlements. Although our model does not necessarily preclude the possibility of settlements, our simplifying assumption does preclude plaintiffs from bringing suits when litigating to judgment is not worthwhile.

be nonpositive for any $x_p \leq \bar{x}$ and positive for any $x_p > \bar{x}$. Section VIB shows how the analysis can be adjusted for the case in which x^* does not equal \bar{x} .

III. FEE-SHIFTING RULES BASED ON THE WINNER'S IDENTITY

In this section, we shall examine the effects of the classic fee-shifting rules, under which fee shifting, to the extent that it occurs, depends only on the identity of the winning party.⁷ Each of the classic fee-shifting rules provides an important benchmark for comparison with fee-shifting rules that can depend on the margin of victory.

A. Judgment Predicted with Certainty

In this section we assume that $x_c = x_p$, so that the plaintiff can predict the judgment with certainty. That is, ϵ is not a random variable. Instead, $\epsilon = 0$ in each case.

The American Rule. When the plaintiff can predict the court's judgment with certainty, the American rule would discourage all frivolous suits but may fail to encourage all meritorious suits. Specifically, under the American rule, the plaintiff will sue if and only if both $x_p > \bar{x}$ and $C_p < D$. A plaintiff will never bring a frivolous suit, because a case with $x_p \leq \bar{x}$ would be bound to lose. Thus, the plaintiff would recover nothing and would be saddled with its own litigation costs. A plaintiff might fail to bring a meritorious suit, however, if its litigation costs are sufficiently large. Even a winning suit, with $x_p > \bar{x}$, would not be worthwhile if $C_p > D$, because the plaintiff would not recover enough to pay its litigation costs.

The British Rule. When the plaintiff can predict the court's judgment with certainty, the British rule would discourage all frivolous suits and encourage all meritorious suits. Specifically, under the British rule, the plaintiff will sue if and only if $x_p > \bar{x}$. Without uncertainty over the trial outcome, this rule yields the optimal incentives for the plaintiff. A plaintiff will never bring a frivolous suit, because it would be bound to lose. Thus, the plaintiff would recover no damages and bear at least its own litigation costs. A plaintiff will always bring a meritorious suit, because it would be bound to win, and in this case the plaintiff would not have to bear its own litigation costs.

The Pro-Defendant Rule. Under the pro-defendant rule, each litigant pays its own costs if the defendant loses, but the plaintiff pays the defen-

⁷ See Steven Shavell, *Suit, Settlement, and Trial: A Theoretical Analysis under Alternative Methods for the Allocation of Legal Costs*, 11 J. Legal Stud. 55 (1982).

dant's costs if the defendant wins. Like the American rule, the pro-defendant rule would discourage all frivolous suits but may fail to encourage all meritorious suits: under conditions of certainty, the plaintiff will sue if and only if both $x_p > \bar{x}$ and $C_p < D$. The pro-defendant rule is no improvement over the American rule. The pro-defendant rule can penalize plaintiffs that bring losing suits, but under conditions of certainty, plaintiffs will not bring such suits under the American rule anyway.

The Pro-Plaintiff Rule. Under the pro-plaintiff rule, each litigant pays its own costs if the plaintiff loses, but the defendant pays the plaintiff's costs if the plaintiff wins. Like the British rule, the pro-plaintiff rule would both discourage all frivolous suits and encourage all meritorious suits: under conditions of certainty, the plaintiff will sue if and only if $x_p > \bar{x}$. The pro-plaintiff rule is an improvement over the American rule. The problem under the American rule is a failure to encourage all meritorious suits, and the pro-plaintiff rule addresses this problem by relieving the winning plaintiff of its litigation costs.

B. Judgment Predicted with Uncertainty

In this section we assume that $x_c = x_p + \epsilon$, where ϵ is a random variable. Thus, the plaintiff cannot predict the trial outcome with certainty. As we shall see, none of the classic rules can guarantee that the plaintiff will have optimal incentives in all cases. Under each rule, plaintiffs bring some frivolous suits and fail to bring some meritorious suits.

The American Rule. Under the American rule, the plaintiff will sue if and only if

$$-C_p + \text{pr}(x_c > \bar{x} | x_p)D > 0. \quad (1)$$

Substituting $x_p + \epsilon$ for x_c in (1), we find this condition is equivalent to

$$\text{pr}(\epsilon > \bar{x} - x_p) > C_p/D. \quad (2)$$

Thus, if $C_p \geq D$, then (2) cannot hold, and the plaintiff will never sue. If $C_p < D$, however, the plaintiff will sue if and only if x_p is greater than some threshold value, which we shall denote as s^* . This s^* is defined by

$$1 - F(\bar{x} - s^*) = C_p/D,$$

or, equivalently,

$$s^* = \bar{x} - F^{-1}[1 - (C_p/D)]. \quad (3)$$

The plaintiff will have optimal incentives if and only if it is just indifferent about bringing the marginal suit (in which $x_p = \bar{x}$), that is, if $s^* = \bar{x}$. Thus, (3) implies that the American rule will lead to optimal incentives

for the plaintiff if and only if $1 - F(0) = C_p/D$. Recall that $1 - F(0)$ equals θ_p , the probability that the court will view the plaintiff's case more favorably than the plaintiff does, which is therefore also the probability that a case with $x_p = \bar{x}$ would succeed.

As in the case of prediction with certainty, this rule might discourage a plaintiff from bringing a meritorious suit if the plaintiff's litigation costs are sufficiently large relative to the amount at stake. The possibility that even a meritorious suit can lose, however, aggravates this problem. Specifically, if $\theta_p < C_p/D$, then too little litigation results. In these cases, either $C_p \geq D$ or $s^* > \bar{x}$, and the plaintiff will be discouraged from bringing some meritorious suits because C_p would be too large relative to D .

Once we allow for prediction with uncertainty, moreover, it is no longer true that this rule would discourage all frivolous suits. The plaintiff might bring a frivolous suit if the plaintiff's litigation costs are sufficiently small relative to the amount at stake, because even a frivolous suit might prevail. Specifically, if $\theta_p > C_p/D$, then too much litigation results. In these cases, $s^* < \bar{x}$, and the plaintiff would bring some frivolous suits, because the likelihood of favorable judgments in these cases would be sufficiently high.

For example, consider the case in which the court is as likely to err in favor of the defendant as it is to err in favor of the plaintiff: $\theta_d = \theta_p = 1/2$. In this case, if $C_p < 1/2 D$, then there will be too much litigation. If $C_p > 1/2 D$ instead, then there will be too little litigation.

The British Rule. Whereas the British rule ensures optimal incentives to sue under conditions of certainty, it no longer does so under conditions of uncertainty. By the same reasoning we applied to the American rule, we can show that under the British rule and conditions of uncertainty, the plaintiff will sue if and only if x_p is greater than some s^* , which is defined by

$$s^* = \bar{x} - F^{-1}[1 - (C_p + C_d)/(D + C_p + C_d)]. \quad (4)$$

Thus, this rule will lead to optimal incentives for the plaintiff if and only if $\theta_p = (C_p + C_d)/(D + C_p + C_d)$.

This rule might discourage a plaintiff from bringing a meritorious suit if the parties' litigation costs are sufficiently large relative to the amount at stake, because even a meritorious suit might lose. Specifically, if $\theta_p < (C_p + C_d)/(D + C_p + C_d)$, then too little litigation results. In these cases, $s^* > \bar{x}$, and the plaintiff will be discouraged from bringing some meritorious suits because the costs that it would bear in the event that the suit loses, $C_p + C_d$, will be sufficiently large relative to the gain in the event the suit wins, D .

Furthermore, this rule might encourage a plaintiff to bring a frivolous suit if the plaintiff's litigation costs are sufficiently small, because even a frivolous suit might prevail. Specifically, if $\theta_p > (C_p + C_d)/(D + C_p + C_d)$, then too much litigation results. In these cases, $s^* < \bar{x}$, and the plaintiff will bring some frivolous suits, because the likelihood of favorable judgments in these cases would be sufficiently high.⁸

For example, consider the case in which the court is as likely to err in favor of the defendant as it is to err in favor of the plaintiff: $\theta_d = \theta_p = 1/2$. In this case, if $C_p + C_d < D$, then there will be too much litigation. If $C_p + C_d > D$ instead, then there will be too little litigation.

The Pro-Defendant Rule. By similar reasoning, we can show that under the pro-defendant rule, if $C_p \geq D$, then the plaintiff would never sue. If $C_p < D$ instead, however, then the plaintiff will sue if and only if x_p is greater than some s^* , which is defined by

$$s^* = \bar{x} - F^{-1}[1 - (C_p + C_d)/(D + C_d)]. \quad (5)$$

Thus, this rule will lead to optimal incentives for the plaintiff if and only if $\theta_p = (C_p + C_d)/(D + C_d)$. If instead $\theta_p < (C_p + C_d)/(D + C_d)$, then too little litigation results. On the other hand, if $\theta_p > (C_p + C_d)/(D + C_d)$, then too much litigation results.⁹

The Pro-Plaintiff Rule. By similar reasoning, we can show that under the pro-plaintiff rule, the plaintiff will sue if and only if x_p is greater than some s^* , which is defined by

$$s^* = \bar{x} - F^{-1}[1 - C_p/(D + C_p)]. \quad (6)$$

Thus, this rule will lead to optimal incentives for the plaintiff if and only

⁸ Note that once we allow for prediction with uncertainty, it is no longer clear whether the British rule or the American rule offers the plaintiff the greater incentive to sue. That is, between the British rule and the American rule, it is ambiguous which rule implies the lower threshold s^* . Comparing (3) and (4), we find that the incentive is greater under the British rule if and only if

$$C_p/C_d > D/(C_p + C_d).$$

For example, in the case in which the litigation costs of the two parties are equal, $C_p = C_d$, the British rule offers the plaintiff greater incentives to sue than the American rule if the total litigation costs exceed the amount at stake but smaller incentives if the amount at stake exceeds total litigation costs.

⁹ For example, consider the case in which $\theta_p = 1/2$. In this case, if $C_p + C_d < 1/2(D + C_d)$, then there will be too much litigation. If $C_p + C_d > 1/2(D + C_d)$ instead, then there will be too little litigation.

If $C_p < D$, then the pro-defendant rule offers the plaintiff less incentive to sue than either the American rule or the British rule: we can see from (5) that the pro-defendant rule implies a threshold s^* that is strictly larger than those under the American and British rules. If $C_p \geq D$ instead, then the incentive to sue would be equally small under the American rule and under the pro-defendant rule: in this case, the plaintiff would never sue under either rule.

if $\theta_p = C_p/(D + C_p)$. If instead $\theta_p < C_p/(D + C_p)$, then too little litigation results. On the other hand, if $\theta_p > C_p/(D + C_p)$, then too much litigation results.¹⁰

IV. THE OPTIMAL ONE-SIDED FEE-SHIFTING RULE

Until now we have examined rules under which, if there is fee shifting at all, it can only depend on the identity of the winner, that is, on whether $x_c > \bar{x}$. In this section we drop this constraint on the structure of the fee-shifting rules. We allow the courts to take into account the margin of victory, that is, we allow the fee-shifting rule to be based on the difference between x_c and \bar{x} .

In this section, we will limit ourselves to one-sided fee-shifting rules, under which, for a given D , $F(\epsilon)$, C_p , and C_d , there may be fee shifting only in favor of one side, throughout the domain of x_c . We will consider two-sided fee shifting in the next section. As we will show, to design the best one-sided fee-shifting rule, it is helpful to distinguish between situations in which the American rule provides inadequate incentives to sue and those in which it provides excessive incentives to sue.

A. *Insufficient Incentive to Litigate under the American Rule*

If $\theta_p < C_p/D$, then the American rule leads to insufficient incentives to bring suit. Under these circumstances, the American rule discourages all frivolous suits but also discourages some meritorious suits. In this case, pro-plaintiff fee shifting can improve the plaintiff's incentives to sue:

PROPOSITION 1. If $\theta_p < C_p/D$, then the best one-sided fee-shifting rule would specify that the defendant pays the plaintiff's litigation costs if $x_c > \bar{y}_p$, where

$$\bar{y}_p = \bar{x} + F^{-1}[1 - (C_p - \theta_p D)/C_p]. \quad (7)$$

This rule would provide optimal results: it would both discourage all frivolous suits and encourage all meritorious suits.

Proof. See the Appendix.

Intuition. Under the American rule, a plaintiff with x_p equal to \bar{x} will expect to win $\theta_p D$ from going to trial at a cost of C_p . If $C_p > \theta_p D$, then

¹⁰ For example, consider the case in which $\theta_p = 1/2$. In this case, if $C_p < 1/2(D + C_p)$, then there will be too much litigation. If $C_p > 1/2(D + C_p)$ instead, then there will be too little litigation.

The incentive to sue is the greatest under the pro-plaintiff rule insofar as (6) implies that the threshold s^* is the lowest under this rule. That is, the pro-plaintiff rule implies a threshold s^* that is strictly smaller than that under any of the other classic fee-shifting rules.

the plaintiff's suit has negative expected value not only for $x_p = \bar{x}$ but also for some x_p slightly greater than \bar{x} . If the plaintiff's litigation costs can be shifted to the defendant with probability close to 1, which we can accomplish with a sufficiently small \bar{y}_p , then the plaintiff can obtain positive expected value from bringing the same cases. Because we can vary \bar{y}_p continuously, we can find an intermediate value for \bar{y}_p such that the expected value of bringing suit will be exactly 0 for the marginal case, in which $x_p = \bar{x}$.

Comparative Statics. Note that the plaintiff's expected payoff from bringing the marginal suit is strictly increasing in θ_p and D but does not depend on C_d . Furthermore, as long as $0 < F(\bar{y}_p - \bar{x}) < 1$, so that fee shifting is possible but not certain in the marginal case, this payoff will strictly decrease in \bar{y}_p and C_p .¹¹ Thus, as long as the fee-shifting rule sets \bar{y}_p to optimize the plaintiff's incentives to bring suit such that $\bar{x} - e_1 < \bar{y}_p < \bar{x} + e_2$, the optimal \bar{y}_p will strictly increase in θ_p and D but strictly decrease in C_p .

That is, *ceteris paribus*, the optimal policy becomes more pro-defendant if either θ_p or D increases but more pro-plaintiff if C_p increases. If the award that the plaintiff expects to recover increases, then the optimal rule—to offset the increased incentive to bring suit—must decrease the probability that the defendant would have to bear the plaintiff's litigation costs. If on the other hand, these litigation costs increase, then the optimal rule—to offset the reduced incentive to bring suit—must decrease the probability that the plaintiff would have to bear them.

Comparison with the Classic Pro-Plaintiff Rule. As we saw in Section IIIB, the classic pro-plaintiff rule, which imposes the constraint $\bar{y}_p = \bar{x}$, is optimal if and only if $\theta_p = C_p/(D + C_p)$. In this section, we see that if $C_p/(D + C_p) < \theta_p < C_p/D$, then the classic pro-plaintiff rule leads to too much litigation, and the best rule instead sets \bar{y}_p higher than \bar{x} . In order to trigger pro-plaintiff fee shifting under the best rule, the plaintiff must not only win its case but also win by a sufficiently wide margin. If $\theta_p < C_p/(D + C_p)$ instead, then the classic pro-plaintiff rule leads to too little litigation, and $\bar{y}_p < \bar{x}$ is necessary to increase the plaintiff's incentives. The best fee-shifting rule in these cases is even more pro-plaintiff than the classic pro-plaintiff fee-shifting rule: the plaintiff wins reimbursement not only when it wins its case but also when it loses its case by a sufficiently small margin.

Fee Shifting in Favor of Winning Plaintiffs Only. Suppose we impose

¹¹ As long as there is some positive probability of $x_c \leq \bar{y}_p$ when $x_p = \bar{x}$, so that the court might leave the burden of C_p on the plaintiff in the marginal suit, then the plaintiff's expected payoff will be strictly decreasing in C_p .

the constraint $\bar{y}_p \geq \bar{x}$, so that the defendant would pay the plaintiff's litigation costs only if the plaintiff prevails. The fee-shifting rule can be no more pro-plaintiff than the classic pro-plaintiff rule. The rule in Proposition 1, however, can offer optimal incentives to the plaintiff consistent with this constraint if and only if $C_p/(C_p + D) \leq \theta_p$. If instead

$$C_p/(C_p + D) > \theta_p, \quad (8)$$

then the constraint will be costly. The best rule consistent with the constraint would be the classic pro-plaintiff rule, under which a plaintiff would sue if and only if $x_p > s^*$, where s^* is defined by equation (6). If condition (8) holds, then equation (6) implies that $s^* > \bar{x}$: plaintiffs that are not sufficiently confident of victory will fail to bring some meritorious suits.

B. Excessive Incentive to Litigate under the American Rule

If $C_p/D < \theta_p$, then the American rule leads to excessive litigation. The American rule encourages all meritorious suits but also encourages some frivolous suits. In this case, pro-defendant fee shifting can improve the plaintiff's incentives to sue:

PROPOSITION 2. If $C_p/D < \theta_p$, then the best one-sided fee-shifting rule would provide for pro-defendant fee shifting. Specifically:

a) If $\theta_p \leq (C_p + C_d)/D$, then the best one-sided fee-shifting rule would specify that the plaintiff pays the defendant's litigation costs if $x_c \leq \bar{y}_d$, where

$$\bar{y}_d = \bar{x} + F^{-1}[1 - (C_p + C_d - \theta_p D)/C_d]. \quad (9)$$

In these cases, this rule would create the optimal incentives for the plaintiff to bring suit.

b) If $\theta_p > (C_p + C_d)/D$, then the best one-sided fee-shifting rule would specify that the plaintiff pays the defendant's litigation costs if $x_c \leq \bar{y}_d$, where $\bar{y}_d \geq \bar{x} + e_2$. This rule would fall short of producing optimal results only insofar as it will fail to discourage all frivolous suits.

Proof. See the Appendix.

Intuition. Note that if litigation costs are sufficiently small relative to the amount at stake, so that

$$\theta_p D > C_p + C_d, \quad (10)$$

then no fee-shifting rule can provide optimal incentives for the plaintiff. In this case, even in the marginal case (in which $x_p = \bar{x}$), the plaintiff's expected payoff from trial would exceed the litigation costs of *both* parties. Even if the plaintiff were certain that it would have to pay the fees

of both parties, these costs would not deter the plaintiff from bringing such a suit. In such cases, fee shifting is an inadequate sanction to deter all frivolous suits.

Comparative Statics. Note that the plaintiff's expected payoff from bringing the marginal suit is strictly increasing in θ_p and D but strictly decreasing in C_p . Furthermore, as long as $0 < F(\bar{y}_d - \bar{x}) < 1$, this payoff will strictly decrease in \bar{y}_d and C_d .¹² Thus, as long as the fee-shifting rule sets \bar{y}_d to optimize the plaintiff's incentives to bring suit such that $\bar{x} - e_1 < \bar{y}_d < \bar{x} + e_2$, the optimal \bar{y}_d will strictly increase in θ_p and D but strictly decrease in C_p and C_d .

That is, *ceteris paribus*, the optimal policy becomes more pro-defendant if either θ_p or D increases but more pro-plaintiff if C_p or C_d increases. If the award that the plaintiff expects to recover increases, then the optimal rule—to offset the increased incentive to bring suit—must increase the probability that the plaintiff would have to bear the defendant's litigation costs. If, on the other hand, either the plaintiff's or the defendant's litigation costs increase, then the optimal rule—to offset the reduced incentive to bring suit—must decrease the probability that the plaintiff would have to bear the defendant's litigation costs.

Comparison with the Classic Pro-Defendant Rule. As we saw in Section IIIB, the classic pro-defendant rule, which imposes the constraint $\bar{y}_d = \bar{x}$, is optimal if and only if $\theta_p = (C_p + C_d)/(D + C_d)$. In this section, we see that if $C_p/D < \theta_p < (C_p + C_d)/(D + C_d)$, then the classic pro-defendant rule leads to too little litigation, and the best rule instead sets \bar{y}_d lower than \bar{x} . In order to trigger pro-defendant fee shifting under the best rule, the defendant must not only win its case but also win by a sufficiently wide margin. If $(C_p + C_d)/(D + C_d) < \theta_p$ instead, then the classic pro-defendant rule leads to too much litigation, and $\bar{y}_d > \bar{x}$ is necessary to reduce the plaintiff's incentives. The best fee-shifting rule in these cases is even more pro-defendant than the classic pro-defendant fee-shifting rule: the defendant wins reimbursement not only when it wins its case, but also when it loses its case by a sufficiently small margin.

Fee Shifting in Favor of Winning Defendants Only. Suppose we impose the constraint $\bar{y}_d \leq \bar{x}$, so that the plaintiff would pay the defendant's litigation costs only if the defendant prevails. The fee-shifting rule can be no more pro-defendant than the classic pro-defendant rule. The rule in Proposition 2, however, can offer optimal incentives to the plaintiff consistent with this constraint if and only if $C_p/D < \theta_p \leq (C_p + C_d)/(D$

¹² As long as there is some positive probability that $x_c \leq \bar{y}_d$ when $x_p = \bar{x}$, so that the court might place the burden of C_d on the plaintiff in the marginal suit, then the plaintiff's expected payoff will be strictly decreasing in C_d .

+ C_d). If instead

$$(C_p + C_d)/(D + C_d) < \theta_p, \quad (11)$$

then the constraint $\bar{y}_d \leq \bar{x}$ will be costly. The best rule consistent with the constraint would be the classic pro-defendant rule, under which a plaintiff would sue if and only if $x_p > s^*$, where s^* is defined by equation (5). If condition (11) holds, then equation (5) implies that $s^* < \bar{x}$: plaintiffs that are not sufficiently convinced of defeat will bring some frivolous suits.

C. Example with a Uniform Distribution

A simple example provides a useful illustration of our results. Suppose for simplicity that ϵ is uniformly distributed in the interval $(-e, e)$, for some $e > 0$. In this special case,

$$F(\epsilon) = 1/2(1 + \epsilon/e), \quad (12)$$

and $\theta_p = 1/2$. Solving (12) for the inverse function, we find

$$F^{-1}(p) = (2p - 1)e \quad (13)$$

for p in the interval $(0, 1)$. Using this particular inverse function in (7) and (9), we can derive the optimal fee-shifting rules.

Proposition 2(a) implies that if $C_p/D < 1/2 < (C_p + C_d)/D$, then the optimal rule specifies that the plaintiff pays the defendant's litigation costs if $x_c \leq \bar{y}_d$, where

$$\bar{y}_d = \bar{x} + (D - 2C_p - C_d)e/C_d. \quad (14)$$

This rule operates only against losing plaintiffs (that is, $\bar{y}_d \leq \bar{x}$) if and only if $D \leq 2C_p + C_d$. Note that \bar{y}_d is increasing in D but decreasing in C_p and C_d . Proposition 2(b) implies that if $(C_p + C_d)/D < 1/2$ instead, then $\bar{y}_d \geq \bar{x} + e$. Finally, Proposition 1 implies that if $1/2 < C_p/D$ instead, then the optimal rule would specify that the defendant pays the plaintiff's litigation costs if $x_c < \bar{y}_p$, where

$$\bar{y}_p = \bar{x} + (D - C_p)e/C_p. \quad (15)$$

This rule operates only against losing defendants (that is, $\bar{y}_p \geq \bar{x}$) if and only if $C_p \leq D$. Note that \bar{y}_p is increasing in D but decreasing in C_p .

V. THE FAMILY OF OPTIMAL FEE-SHIFTING RULES

In this section, we examine the possibility of fee-shifting rules under which, even for a given D , $F(\epsilon)$, C_p , and C_d , the court may require either side to reimburse the other for its litigation costs. That is, we do not

require either $\bar{y}_p = \infty$ or $\bar{y}_d = -\infty$ to hold; instead, we can allow both \bar{y}_p and \bar{y}_d to be finite for the same set of cases. We will consider how to design the best two-sided fee-shifting rule. We analyze the best two-sided rule to see if it performs any better than the best one-sided rule, and as we will see, it does not. We also study two-sided fee-shifting rules because Rule 11 is an example of such a rule. Our analysis of how best to design such a rule will shed light on the use of fee shifting under Rule 11.

It will be helpful in this analysis to distinguish between the case in which $\theta_p \leq (C_p + C_d)/D$ and the case in which $\theta_p > (C_p + C_d)/D$. In the first case, fee shifting can produce the optimal incentives for plaintiffs to bring suit. In the second case, we will see that it cannot.

A. *The Case in Which Fee Shifting Can Produce Optimal Results*

Let us look first at the case in which condition (10) does not hold: $\theta_p \leq (C_p + C_d)/D$. In this case, if the plaintiff must always bear the litigation costs of both sides, then all frivolous suits would be discouraged. The probability of error in favor of the plaintiff would not be sufficient to induce the plaintiff to bear these costs. Thus, in this case one could conceive of a fee-shifting rule that would deter all frivolous suits. In fact, we find that there exists a whole family of two-sided fee-shifting rules that can ensure that the plaintiff brings suit if and only if the case is sufficiently strong:

PROPOSITION 3. If $\theta_p \leq (C_p + C_d)/D$, then there exists a family of two-sided fee-shifting rules that would create the optimal incentives for the plaintiff to bring suit. Each such rule would specify that the defendant pays the plaintiff's expenses if $x_c > \bar{y}_p$ and that the plaintiff pays the defendant's expenses if $x_c \leq \bar{y}_d$, where

$$-C_p + [1 - F(\bar{y}_p - \bar{x})]C_p - F(\bar{y}_d - \bar{x})C_d + \theta_p D = 0. \quad (16)$$

All of these rules would discourage all frivolous suits and encourage all meritorious suits.

Proof. See the Appendix.

Intuition. As long as condition (10) does not hold, there is a one-sided fee-shifting rule that would produce optimal incentives for plaintiffs to bring suit. The best two-sided fee-shifting rule can produce no better results, and as Proposition 3 shows, it produces no worse results. We can design a two-sided fee-shifting rule, indeed a whole family of such rules, that would produce optimal results.

The optimal two-sided fee-shifting rule, like the optimal one-sided fee-shifting rule, ensures that the expected value of going to trial is negative

for $x_p < \bar{x}$, zero for the marginal suit in which $x_p = \bar{x}$, and positive for $x_p > \bar{x}$. Without fee shifting, the expected value of the marginal suit would be $-C_p + \theta_p D$. Fee shifting adds the second and third terms to the expression for the expected value of the marginal suit on the left-hand side of equation (16). Equation (16) sets the expected value of the marginal suit, including the expected net gain (or loss) from fee shifting, equal to 0.

Extreme Thresholds for Fee Shifting. In designing a two-sided fee shifting rule, there would never be any reason to set $\bar{y}_d < \bar{x} - e_1$ or to set $\bar{y}_p > \bar{x} + e_2$. Consider a rule with $\bar{y}_d < \bar{x} - e_1$: there would be no reason not to increase \bar{y}_d to $\bar{x} - e_1$, because this change would only increase the expected costs of frivolous suits; it would not affect meritorious suits. Similarly, if we are considering a rule with $\bar{y}_p > \bar{x} + e_2$, then there is no reason not to lower \bar{y}_p to $\bar{x} + e_2$, because this change would only reduce the expected costs of meritorious suits; it would not affect frivolous suits. Neither of these effects would be undesirable. If the other threshold is set optimally, then making either of these changes would be unnecessary, because the rule would already discourage all frivolous suits and encourage all meritorious suits. If the other threshold is not set optimally, however, then these changes might be desirable.

Fee Shifting to Encourage Suits. Suppose $C_p > \theta_p D$, so that there is an insufficient incentive for the plaintiff to bring suit under the American rule. In this case, the optimal one-sided fee-shifting rule provided for pro-plaintiff fee shifting and ruled out any pro-defendant fee shifting. That is, it set $\bar{y}_d = -\infty$. Note that with the optimal two-sided fee-shifting rule, as long as $\bar{y}_d > \bar{x} - e_1$, so that there is some probability of pro-defendant fee shifting in the marginal case, we must increase the probability of pro-plaintiff fee shifting above what it would be under the optimal one-sided fee-shifting rule.

Note also that the sum of the second and third terms in equation (16) must equal $C_p - \theta_p D$. Thus, if $C_p > \theta_p D$, then this sum must be positive. If we suppose further that $C_p = C_d$, then $1 - F(\bar{y}_p - \bar{x})$ must exceed $F(\bar{y}_d - \bar{x})$. Therefore, under these assumptions, we must set the fee-shifting thresholds such that the probability of pro-plaintiff fee shifting must exceed the probability of pro-defendant fee shifting. For example, if $\bar{y}_d < \bar{x} < \bar{y}_p$ under our rule, and the distribution of ϵ is symmetric about 0, then \bar{y}_p must be closer to \bar{x} than \bar{y}_d is.

Fee Shifting to Discourage Suits. Suppose $C_p < \theta_p D$, so that there is an excessive incentive for the plaintiff to bring suit under the American rule. In this case, the optimal one-sided fee-shifting rule provided for pro-defendant fee shifting and ruled out any pro-plaintiff fee shifting. That is, it set $\bar{y}_p = \infty$. Note that with the optimal two-sided fee-shifting rule,

as long as $\bar{y}_p < \bar{x} + e_2$, so that there is some probability of pro-plaintiff fee shifting in the marginal case, we must increase the probability of pro-defendant fee shifting above what it would be under the optimal one-sided fee-shifting rule.

Note also that if $C_p < \theta_p D$, then the sum of the second and third terms in equation (16) must be negative. If we suppose further that $C_p = C_d$, then $F(\bar{y}_d - \bar{x})$ must exceed $1 - F(\bar{y}_p - \bar{x})$. Therefore, under these assumptions, we must set the fee-shifting thresholds such that the probability of pro-defendant fee shifting exceeds the probability of pro-plaintiff fee shifting. For example, if $\bar{y}_d < \bar{x} < \bar{y}_p$ under our rule, and the distribution of ϵ is symmetric about 0, then \bar{y}_d must be closer to \bar{x} than \bar{y}_p is.

Comparative Statics. Note that the plaintiff's expected payoff from bringing the marginal suit is monotonically increasing in θ_p and D . Furthermore, as long as both $0 < F(\bar{y}_d - \bar{x}) < 1$ and $0 < F(\bar{y}_p - \bar{x}) < 1$, this payoff will strictly decrease in \bar{y}_d , \bar{y}_p , C_d , and C_p .¹³ Thus, if the fee-shifting rule sets both thresholds, \bar{y}_d and \bar{y}_p , to optimize the plaintiff's incentives to bring suit such that $\bar{x} - e_1 < \bar{y}_d < \bar{x} + e_2$ and $\bar{x} - e_1 < \bar{y}_p < \bar{x} + e_2$, then each threshold—holding the other constant—will increase in θ_p and D but decrease in C_d and C_p .

That is, *ceteris paribus*, the optimal policy becomes more pro-defendant if either θ_p or D increases but more pro-plaintiff if either C_d or C_p increases. If the award that the plaintiff expects to recover increases, then the optimal rule—to offset the increased incentive to bring suit—must increase the probability that the plaintiff would have to bear the parties' litigation costs. If, on the other hand, these litigation costs increase, then the optimal rule—to offset the reduced incentive to bring suit—must decrease the probability that the plaintiff would have to bear them.

B. The Case in Which Fee Shifting Cannot Produce Optimal Results

Consider the case in which condition (10) holds. If $\theta_p > (C_p + C_d)/D$, then we cannot discourage all frivolous suits, even if the plaintiff must always bear the litigation costs of both sides. The probability of error in favor of the plaintiff would be sufficient to induce the plaintiff to bear these costs in the marginal case: the plaintiff still anticipates positive

¹³ As long as the court threatens to impose both C_d and C_p on the plaintiff in the marginal suit, its payoff will decrease in both parameters. In particular, if there is some positive probability of $x_c \leq \bar{y}_d$ when $x_p = \bar{x}$, then this payoff will be strictly decreasing in C_d . Similarly, if there is some positive probability of $x_c \leq \bar{y}_p$ when $x_p = \bar{x}$, then this payoff will be strictly decreasing in C_p also.

expected value from going to trial. Thus, in this case there is no fee-shifting rule that would deter all frivolous suits.

PROPOSITION 4. If $\theta_p > (C_p + C_d)/D$, then the best two-sided fee-shifting rule would specify that the plaintiff pays the defendant's expenses if $x_c \leq \bar{y}_d$ and that the defendant pays the plaintiff's expenses if $x_c > \bar{y}_p$, where $\bar{y}_d \geq \bar{x} + e_2$ and $\bar{y}_p \geq \bar{x} + e_2$. Under this rule, however, the plaintiff would still have excessive incentives to bring suit.

Remark. If condition (10) holds, then the best one-sided fee-shifting rule would set \bar{y}_d to ensure that the plaintiff with the marginal case (or any worse case) would always bear the defendant's litigation costs. The best two-sided fee-shifting rule does the same and also sets \bar{y}_p such that the defendant would never have to pay the plaintiff's litigation costs in a marginal case (or in any worse case). This rule would be the best two-sided fee-shifting rule in that it maximizes the achievement of each objective: this rule would encourage all meritorious suits and discourage as many frivolous suits as possible. This rule would fall short of producing optimal results insofar as it will fail to discourage all frivolous suits.

C. Fee Shifting Subject to Constraints on \bar{y}_p and \bar{y}_d

Each of the classic fee-shifting rules is a special restricted case of the more general two-sided fee-shifting rule, with (\bar{y}_d, \bar{y}_p) constrained to take on a particular pair of values: $(-\infty, \infty)$ for the American rule; (\bar{x}, \bar{x}) for the British rule; $(-\infty, \bar{x})$ for the pro-plaintiff rule; and (\bar{x}, ∞) for the pro-defendant rule. As we have seen, these constrained rules provide optimal incentives only in very special circumstances. Similarly, constraints such as $\bar{y}_d \leq \bar{x}$ or $\bar{y}_p \geq \bar{x}$ would also restrict the circumstances under which the two-sided fee-shifting rule could provide optimal incentives. These restrictions, however, prove to be less severe than those imposed by the classic fee-shifting rules.

The constraint $\bar{y}_p \geq \bar{x}$, for example, would imply that the fee-shifting rule can be no more pro-plaintiff than the classic pro-plaintiff rule. Recall that even the classic pro-plaintiff rule leads to too little litigation if and only if condition (8) holds. Thus, fee shifting subject to the constraint $\bar{y}_p \geq \bar{x}$ can provide optimal incentives if and only if the opposite condition holds, which we can express as

$$C_p \theta_d \leq \theta_p D. \quad (17)$$

Condition (17) states that in the marginal case, the plaintiff's expected gain from adjudication must at least equal its expected liability for its litigation costs under the pro-plaintiff rule.

If the plaintiff's litigation costs are too great relative to the amount at

stake, then (8) rather than (17) holds. Then even under the pro-plaintiff rule, plaintiffs sue if and only if $x_p > s^*$, where s^* is defined by equation (6). As discussed in Section IVA, condition (8) implies $s^* > \bar{x}$, so that plaintiffs fail to bring some meritorious cases because they are sufficiently close cases and the risk of losing is too great.

Similarly, the constraint $\bar{y}_d \leq \bar{x}$ would imply that the fee-shifting rule can be no more pro-defendant than the classic pro-defendant rule. Recall that even the classic pro-defendant rule leads to too much litigation if and only if condition (11) holds. Thus, fee shifting subject to the constraint $\bar{y}_d \leq \bar{x}$ can provide optimal incentives if and only if the opposite condition holds, which we can express as

$$\theta_p D \leq C_p + C_d \theta_d. \quad (18)$$

Condition (18) states that in the marginal case, the plaintiff's expected gain from adjudication cannot exceed its expected liability for litigation costs under the pro-defendant rule.

If litigation costs are too small relative to the amount at stake, then (11) rather than (18) holds. Then as discussed in Section IVB, condition (11) implies that the s^* defined by equation (5) is less than \bar{x} . Thus, even under the pro-defendant rule, plaintiffs bring some frivolous suits where $x_p > s^*$, because they are sufficiently close cases and the possibility of winning is great enough to make the effort worthwhile.

D. Example with a Uniform Distribution

Suppose again for simplicity that ϵ is uniformly distributed in the interval $(-e, e)$. As long as condition (10) does not hold, we can use the $F(\epsilon)$ in (12) to derive the family of optimal rules for this example. After rearranging terms, we find that Proposition 3 and (12) together imply that if $1/2 \leq (C_p + C_d)/D$, then the optimal rule ensures that

$$\bar{y}_p C_p + \bar{y}_d C_d = (D - C_p - C_d)e + (C_p + C_d)\bar{x}, \quad (19)$$

where \bar{y}_p and \bar{y}_d both lie in the interval $(\bar{x} - e, \bar{x} + e)$.¹⁴ Solving (19) for \bar{y}_p yields

$$\bar{y}_p = \bar{x} + [(D - C_p - C_d)e - (\bar{y}_d - \bar{x})C_d]/C_p, \quad (20)$$

and solving (19) for \bar{y}_d yields

$$\bar{y}_d = \bar{x} + [(D - C_p - C_d)e - (\bar{y}_p - \bar{x})C_p]/C_d. \quad (21)$$

Note that this example reveals another comparative statics result: the effect of an increase in the uncertainty of the trial outcome has an ambiguo-

¹⁴ Proposition 4 implies that if $(C_p + C_d)/D < 1/2$ instead, then $\bar{y}_d \geq \bar{x} + e$ and $\bar{y}_p \geq \bar{x} + e$.

ous effect on the (\bar{y}_d, \bar{y}_p) locus. An increase in e may move the locus to higher values of \bar{y}_p and \bar{y}_d or to lower values, depending on whether D is greater or less than $C_p + C_d$.

VI. EXTENSIONS

A. Plaintiff Uncertainty Regarding x

An important element in the analysis is the presence of legal uncertainty: the plaintiff may be uncertain about the outcome of a trial. One source of such uncertainty, and the one we have used for concreteness in our model, is the possibility of an erroneous decision by the court: even if the plaintiff can observe the "true merit" of its case, the court might not observe the "true merit" accurately, and the plaintiff cannot completely predict the court's error. A second possible source of uncertainty is the possibility of an erroneous evaluation by the plaintiff: the plaintiff might be uncertain about the "true merit" of its own case.

Until now we have not considered this second possibility: we have assumed that $x_p = x$. We can assume instead that not only the court but also the plaintiff observes x , the "true merit" of the case, with error. Specifically, let $x_p = x + \epsilon_p$ and $x_c = x + \epsilon_c$, with both ϵ_p and ϵ_c as random variables. In this case, $x_c = x_p + \epsilon$, as before, where now $\epsilon = \epsilon_c - \epsilon_p$. The model yields the same results as before, with the random variable ϵ now reinterpreted to represent the difference between the court's error and the plaintiff's error.

As this discussion suggests, in fact, it is not important for our results whether any "true" value of x exists at all. Although we have assumed that there is a "true" value for x , we can also give our framework an interpretation under which there is no "true" x . Instead, there is only what the court will determine x to be, x_c , and the plaintiff's expectation as to what the court's determination will be. In this case, let x_p denote the expected value attached by the plaintiff to x_c . The plaintiff knows only that $x_c = x_p + \epsilon$, where ϵ is distributed according to $F(\epsilon)$, and the analysis proceeds as before.

B. Other Social Objectives: Different Thresholds for Desirable Suits

In the preceding analysis, we have assumed that x^* , the threshold that x_p must exceed for a suit to be socially desirable, equals \bar{x} . That is, we took the policy goal to be to induce the plaintiff to sue if and only if it observes x_p greater than \bar{x} . We can, however, allow for x^* other than \bar{x} .

A threshold different from \bar{x} is especially plausible once we consider plaintiffs that observe x only with error. Suppose that $x_p = x + \epsilon_p$, where

ϵ_p is distributed symmetrically about 0. In this case, setting $x^* = \bar{x}$ implies that plaintiffs should sue if and only if they believe that the probability of $x > \bar{x}$ is at least 50 percent. One can just as easily consider a policy that would require plaintiffs to sue with a confidence level greater than 50 percent or a confidence level lower than 50 percent.

For example, courts may identify a class of suits that society would not want to discourage even though the plaintiffs regard their claims as unlikely to succeed. There may be some public good flowing from these suits, even if plaintiffs estimate that their likelihood of success is substantially less than 50 percent. A suit may be unlikely to succeed, for example, because it relies on innovative legal theories that promote the development of legal doctrines. In terms of our model, if a suit is likely to fall short of success on this particular ground, courts may set $x^* < \bar{x}$ to encourage such litigation. That is, to the extent that any other social objective militates in favor of plaintiffs bringing some particular types of lawsuits, a more pro-plaintiff policy would be appropriate.¹⁵

This extension requires only minor modifications in the preceding analysis. Again, we would set \bar{y}_p or \bar{y}_d (or both) so that the expected value for the plaintiff of going to trial with the marginal suit is 0, but now we regard the case in which $x_p = x^*$ (rather than $x_p = \bar{x}$) to be the marginal suit. Thus, x^* would replace \bar{x} where appropriate. The plaintiff is still indifferent about bringing the marginal suit, but we have changed the probability of success of the marginal suit. Thus, such an extension would still preserve the general thrust of our results.

A lower x^* implies a lower expected value for the marginal suit in the absence of fee shifting. If we reduce x^* , then we are more likely to

¹⁵ The optimal threshold x^* should ultimately be derived in a comprehensive normative analysis that would consider how the magnitude and the allocation of litigation costs would affect the defendant's incentives *ex ante*. See Keith N. Hylton, *The Influence of Litigation Costs on Deterrence under Strict Liability and under Negligence*, 10 *Int'l Rev. L. & Econ.* 161 (1990); Keith N. Hylton, *Litigation Cost Allocation Rules and Compliance with the Negligence Standard*, 22 *J. Legal Stud.* 457 (1993); Ivan P. L. P'ng, *Litigation, Liability, and Incentives for Care*, 34 *J. Pub. Econ.* 61 (1987). This analysis would also seek to reduce total litigation costs. See Polinsky & Rubinfeld, *Sanctioning Frivolous Suits*, *supra* note 3; Polinsky & Rubinfeld, *Optimal Awards*, *supra* note 3. The goal of optimal fee-shifting rules would be to align the plaintiff's private incentives to bring a suit with the net social benefits of that suit. See Steven Shavell, *The Social versus Private Incentive to Bring Suit in a Costly Legal System*, 11 *J. Legal Stud.* 333 (1982). Nevertheless, because stronger cases would be more likely to contribute to our social objectives than weaker cases, it seems plausible in any event that plaintiffs with sufficiently strong cases should sue whereas plaintiffs with sufficiently weak cases should not. Therefore, we believe that our framework, which determines how to induce plaintiffs to sue if and only if they believe their cases to be sufficiently strong, is of general interest.

have an inadequate incentive to sue (and less likely to have an excessive incentive to sue) in the absence of fee shifting. Thus, a reduction in x^* implies that the best fee-shifting rule is more pro-plaintiff: it is more likely to put the burden of litigation costs on the defendant, with a lower \bar{y}_p or \bar{y}_d (or a set of lower \bar{y}_p and \bar{y}_d values).¹⁶

C. Policy Instruments Other than Fee Shifting

Thus far, we have assumed that the court can impose only fee shifting as a sanction. The court simply sets the thresholds that will trigger fee shifting. More generally, as already suggested, one could also make the size of the sanction a policy variable. That is, one could imagine rules under which the defendant pays the plaintiff S_p if $x_c > \bar{y}_p$ and the plaintiff pays the defendant S_d if $x_c \leq \bar{y}_d$. Even more generally, you could have the net transfer S from defendant to plaintiff be a nondecreasing function of x_c . In that case, the plaintiff's expected payoff from litigation still increases monotonically in x_p . The basic approach of this article still applies: the court can set the function $S(x_c)$ such that the plaintiff is just indifferent about bringing the marginal suit but willing to bring better suits and unwilling to bring worse suits.

Given that the absolute value of $S(x_c)$ can be raised over the level of reimbursement for litigation costs, it should always be possible to induce optimal decisions, even if the court can impose sanctions only on the losing party. For example, in the case discussed in Section VC in which condition (17) fails to hold, we would have to supplement or replace fee shifting subject to the constraint $\bar{y}_p \geq \bar{x}$ with other policies in order to induce plaintiffs to bring suit in all appropriate cases. Courts might provide for additional awards in some cases to induce plaintiffs to bring meritorious suits that they would otherwise not bring because of high litigation costs. That is, we can increase the expected value of the plaintiff's suit by shifting the $S(x_c)$ function up.¹⁷

If a court may impose sanctions paid by the defendant to the plaintiff in excess of the plaintiff's attorney's fees, then courts could use this

¹⁶ Conversely, a higher x^* implies a more pro-defendant rule. A higher x^* implies a higher expected value for the marginal suit in the absence of fee shifting. If we raise x^* , then we are more likely to have an excessive incentive to sue (and less likely to have an inadequate incentive to sue) in the absence of fee shifting. Thus, an increase in x^* implies that the best fee-shifting rule is less pro-plaintiff: it is more likely to put the burden of litigation costs on the plaintiff, with a higher \bar{y}_p or \bar{y}_d (or a set of higher \bar{y}_p and \bar{y}_d values).

¹⁷ Similarly, if condition (18) fails to hold, then we would have to supplement or replace fee shifting subject to the constraint $\bar{y}_d \leq \bar{x}$ with other policies in order to discourage plaintiffs from bringing any frivolous suits. Specifically, we can decrease the expected value of the plaintiff's suit by shifting the $S(x_c)$ function down.

authority to encourage plaintiffs to bring suit. There are also other policy instruments, however, that can shift the $S(x_c)$ function. Suppose for example that a court is more likely to award punitive damages in a tort case if it finds that the defendant was grossly negligent. In this case, the expected damages anticipated by the plaintiff will depend on the plaintiff's observed x_p , because the amount awarded upon a finding of liability will be correlated with the court's findings on x . By increasing the punitive damages or awarding them more readily, courts could shift the $S(x_c)$ function up. Or we could provide for treble damages in a particular category of cases to increase the amount recovered, so that condition (17) can hold.

VII. IMPLICATIONS FOR RULE 11

Federal Rule of Civil Procedure 11 requires an attorney (or a party not represented by an attorney) to sign pleadings, motions, and other papers before filing them in court, thereby certifying that to the best of that person's knowledge after a reasonable inquiry the paper has not been presented "for any improper purpose, such as to harass or to cause unnecessary delay or needless increase in the cost of litigation," is "warranted by existing law or by a nonfrivolous argument for the extension, modification, or reversal of existing law," and is well grounded in fact.¹⁸ If a court determines that an attorney or party has violated this rule, the court, on motion or on its own initiative, shall impose on that person "an appropriate sanction," which may include an order to pay to the other party or parties the amount of the expenses incurred as a result of the violation, including reasonable attorneys' fees.

For example, if the plaintiff's case is so frivolous that the court finds that the plaintiff's attorney should have known the suit was without merit when filed, the rule allows the court to shift the burden of the defendant's fees to the plaintiff. In enforcing Rule 11, courts have often focused on the merits of claims and defenses in this way. Furthermore, in the overwhelming majority of cases imposing sanctions under Rule 11, courts have punished the party filing the "frivolous" paper by awarding costs and fees to the opposing party.¹⁹ Thus, the Court of Appeals for the Seventh Circuit has stated that "Rule 11 is a fee-shifting statute."²⁰ Crit-

¹⁸ Amendments to the Federal Rules of Civil Procedure and Forms 16-17 (transmitted to Congress on April 22, 1993) (hereinafter 1993 Amendments), reprinted in 146 F.R.D. 405, 420-21 (1993).

¹⁹ See Melissa L. Nelken, Sanctions under Amended Federal Rule 11—Some "Chilling" Problems in the Struggle between Compensation and Punishment, 74 Geo. L. J. 1313, 1333 (1986).

²⁰ Hays v. Sony Corp., 847 F.2d 412, 419 (7th Cir. 1988).

ics of this interpretation of Rule 11, however, warn against routine use of expense shifting as a sanction.²¹ In response to this criticism, the authors of the 1993 amendments modified the rule so that Rule 11(c)(2) now states that the court may order payment of "some or all of the reasonable attorneys' fees and other expenses incurred as a direct result of the violation" if such cost-shifting is "warranted for effective deterrence."²²

Our analysis sheds light on how courts can best use Rule 11 to deter frivolous suits while encouraging meritorious suits. Although courts commonly use fee shifting as a Rule 11 sanction, they can also use other sanctions. We will first assume that fee shifting is the Rule 11 sanction and examine the determination of the appropriate thresholds for fee shifting. We will then comment on circumstances in which fee shifting will be inadequate and other sanctions are necessary.

A. *The Threshold for Fee Shifting*

Let us suppose that courts will use fee shifting as the Rule 11 sanction. Within our framework, Rule 11 is a two-sided fee-shifting rule, because it allows for the possibility of fee shifting in either direction in each case. Our analysis sheds light on how the thresholds for fee shifting should be

²¹ Burbank, for example, argues that courts should exercise greater discretion in selecting a sanction sufficient to deter in the particular circumstances of the case. See Stephen B. Burbank, *The Transformation of American Civil Procedure: The Example of Rule 11*, 137 U. Pa. L. Rev. 1925 (1989).

²² 1993 Amendments at 19, reprinted in 146 F.R.D. at 423. The authors of the 1993 amendments to Rule 11 concluded that the rule had "too rarely been enforced through nonmonetary sanctions" and "cost-shifting . . . has too frequently been selected as the sanction." Attachment B to letter from Sam C. Pointer, Jr., Chairman, Advisory Committee on Civil Rules to Robert E. Keeton, Chairman, Standing Committee on Rules of Practice and Procedure 3-4 (May 1, 1992) (hereinafter Attachment B), reprinted in 146 F.R.D. 521, 523-24 (1993). Consequently, the authors sought to emphasize deterrence as the goal of Rule 11 and to discourage reliance on monetary sanctions, especially fee shifting.

The model in this article, however, indicates that tailoring the deterrent effect of Rule 11 sanctions is more complicated given the existence of uncertainty over trial outcomes. If courts cannot observe x_p , then they cannot know how likely the plaintiff thought its suit was to succeed at trial. Given this uncertainty, a court cannot tailor the sanction in any particular case to that level just sufficient to deter that particular plaintiff. The deterrent effect on any particular plaintiff is also a function of all the other sanctions that the plaintiff considered possible. That is, this deterrent effect depends not only on the sanction imposed in any particular case, as the court views the case, but also on what sanctions the court would impose if it viewed the case differently—that is, if it observed a different x_c .

Viewed in this light, fee shifting can be more flexible and more useful "for effective deterrence" than implied by the Advisory Committee. Our analysis suggests how courts can control Rule 11's deterrent effect not only by varying the magnitude of the sanctions but also by varying the thresholds that will trigger the sanctions. That is, this article suggests that courts can tailor the deterrent effect to the particular circumstances of the case by choosing the conditions under which sanctions would be imposed, not necessarily by adjusting the severity of the sanctions.

set. Given particular definitions of frivolous and meritorious suits, we have shown how the optimal thresholds would depend on the various parameters involved.

Given enough information on $F(\epsilon)$, for example, there may be cases where fee shifting would obviously be appropriate. Whenever the plaintiff wins by such a large margin that there is no chance that the plaintiff has brought a frivolous suit but won through judicial error, then the defendant should pay the plaintiff's litigation costs. This fee-shifting policy might encourage meritorious suits that would not otherwise be brought, and it would not encourage any frivolous suits. Similarly, whenever the defendant wins by such a large margin that there is no chance that the plaintiff has brought a meritorious suit but lost through judicial error, then the plaintiff should pay the defendant's litigation costs. This fee-shifting policy might discourage such frivolous suits, and it would not discourage any meritorious suits.

We have seen, however, that it may well be desirable to shift fees in a larger set of cases. In particular, a court might inquire whether there is, in the absence of such fee shifting, either insufficient or excessive incentives for the plaintiff to bring suit. To this end, the court would need to determine the expected value for the plaintiff of going to trial with the marginal case. This expected value depends on the plaintiff's litigation costs, the amount at stake, and the likelihood of judicial error on the merits in either direction.²³ If this expected value is negative, then the plaintiff has insufficient incentives to sue; if this expected value is positive, then the plaintiff has excessive incentives to sue.

If there is an insufficient incentive to sue in the absence of fee shifting, then the ideal fee-shifting rule would offer the plaintiff a net expected gain in the marginal case. That is, the thresholds should be set in such a way that the expected pro-plaintiff fee shifting would exceed the expected pro-defendant fee shifting. If the plaintiff's and defendant's litigation costs are similar, then the thresholds should imply that the likelihood of pro-plaintiff fee shifting exceeds the likelihood of pro-defendant fee shifting.²⁴

²³ To set precisely the best thresholds for fee shifting requires more information than courts are likely to have in reality. Nevertheless, our analysis suggests at least some crude rules of thumb. Although courts would not be able to implement perfectly the best fee-shifting rule, they should be able to achieve roughly the desired effect on plaintiffs' incentives. They may shift fees in some cases in which the best rule would not and fail to shift fees in some cases in which the best rule would call for it. Unless one type of departure from the best rule predominates, however, the two types of error in fee-shifting decisions would tend to offset one another. Such random errors—without systematic bias—would have little net effect on the plaintiff's expected value from bringing suit.

²⁴ Thus, the thresholds should be set so as to favor the plaintiff: if errors are distributed symmetrically about zero and litigation costs are similar, then the threshold for pro-plaintiff

In contrast, if there is an excessive incentive to sue in the absence of fee shifting, then the ideal fee-shifting rule would impose a net expected cost on the plaintiff in the marginal case. That is, the thresholds should be set in such a way that the expected pro-defendant fee shifting would exceed the expected pro-plaintiff fee shifting. If the plaintiff's and defendant's litigation costs are similar, then the thresholds should imply that the likelihood of pro-defendant fee shifting exceeds the likelihood of pro-plaintiff fee shifting.²⁵

In any case, the comparative statics results in Section VA offer some further guidance for courts exercising the discretion that they enjoy under Rule 11. These results suggest that courts should give Rule 11 an interpretation more generous to plaintiffs in cases in which the litigation costs are larger relative to the amount at stake. As Section VA indicates, this interpretation may mean that courts use Rule 11 more sparingly against plaintiffs, or that they use it more aggressively against defendants, or both. Conversely, if the amount at stake is larger relative to the litigation costs, then courts should invoke Rule 11 more readily against plaintiffs and less frequently against defendants.

We have assumed that the plaintiff is risk neutral. Given a risk-averse plaintiff, an increase in risk would magnify the deterrent effect of any given fee-shifting rule. This effect suggests that courts should give Rule 11 a more pro-plaintiff interpretation in the face of greater legal uncertainty or when dealing with particularly risk-averse plaintiffs. Thus, it may be appropriate for courts to afford more favorable treatment under Rule 11 to plaintiffs that bring suit in unsettled areas of the law or those with lower levels of wealth because more pro-plaintiff thresholds would be necessary to achieve a given deterrent effect.²⁶

We may also want to vary our fee-shifting thresholds because we wish to deter some suits more than others even if they are equally unlikely to win on the merits.²⁷ That is, we may believe some suits are socially

fee shifting should be closer to the standard for liability than the threshold for pro-defendant fee shifting is.

²⁵ Thus, the thresholds should be set so as to favor the defendant: if errors are distributed symmetrically about zero and litigation costs are similar, then the threshold for pro-defendant fee shifting should be closer to the standard for liability than the threshold for pro-plaintiff fee shifting is.

²⁶ To put this suggestion in the language of Rule 11, in unsettled areas of the law, more lawsuits will be "warranted by existing law or by a nonfrivolous argument for the extension, modification, or reversal of existing law." Furthermore, the Advisory Committee identified "cases involving litigants with greatly disparate financial resources" as examples of cases "in which cost-shifting may be needed for effective deterrence." Attachment B at 4, reprinted in 146 F.R.D. at 524. The committee presumably intended to endorse cost shifting in favor of the party with fewer resources.

²⁷ For example, fee shifting might be more pro-defendant if the plaintiff's attorney failed to conduct a reasonable prefiling inquiry into the merits of the suit. Schwarzer argues that

desirable even if the plaintiff has a relatively low probability of winning. If we wish plaintiffs to bring such suits even when they have little confidence of success, then we would need to set more pro-plaintiff thresholds in such cases. Controversy over Rule 11 focuses on the risk that sanctions will discourage the use of innovative legal theories that produce important social benefits.²⁸ Courts can respond to these concerns by allowing an attorney greater leeway in making a losing argument when it is a novel "argument for the extension, modification, or reversal of existing law," in the language of Rule 11.²⁹ The optimal thresholds will depend in this way on the social benefits and costs of the type of suit in question.³⁰

courts should shift their scrutiny in Rule 11 cases from the merits of the case to the adequacy of the attorney's prefiling inquiry. See William W. Schwarzer, Rule 11 Revisited, 101 Harv. L. Rev. 1013 (1988). Schwarzer argues that such a shift from the merits to the attorney's conduct would promote more predictable and less costly enforcement of Rule 11, because it is difficult to predict whether a court will find a particular paper "frivolous." See Sanford Levinson, Frivolous Cases: Do Lawyers Really Know Anything at All? 24 Osgoode Hall L. J. 353 (1986). As we have seen, however, attorneys do not need to predict sanctions with certainty for Rule 11 to exert the desired deterrent effect. Instead, in our framework, we might set more demanding thresholds for a plaintiff bringing suit without conducting an adequate inquiry (than we do for a plaintiff filing the same suit after an inquiry), even though such a derelict plaintiff would predict the trial outcome with greater error, simply because we wish to deter such behavior. The prefiling inquiry is socially desirable because it might have revealed that the case was without merit.

²⁸ See Nelken, *supra* note 19; Note, Plausible Pleadings: Developing Standards for Rule 11 Sanctions, 100 Harv. L. Rev. 630 (1987).

²⁹ Wilder argues that courts have in fact been sensitive to the chilling effect of sanctions and have used Rule 11 cautiously. See Nancy H. Wilder, The 1983 Amendments to Rule 11: Answering the Critics' Concern with Judicial Self-Restraint, 61 Notre Dame L. Rev. 798 (1986).

³⁰ The optimal thresholds would also depend on the costs of fee shifting itself. Fee shifting may entail additional costs, for example, if parties respond by expending greater resources litigating the merits of the case. See Ronald Braeutigam, Bruce Owen & John Panzar, An Economic Analysis of Alternative Fee Shifting Systems, 47 L. & Contemp. Probs. 173 (1984); John C. Hause, Indemnity, Settlement, and Litigation, or, I'll Be Suing You, 18 J. Legal Stud. 157 (1989). Furthermore, fee shifting depends on the resolution of issues other than the court's determination on the merits. In our framework, a court must determine $F(e)$, C_d , and C_p , as well as the margin of victory, and in practice these issues might be costly for the parties and the court to resolve. In fact, critics have often charged that Rule 11 has generated too much costly "satellite litigation." See A. Leon Higginbotham, Jr. *et al.*, Bench-Bar Proposal to Revise Civil Procedure Rule 11, 137 F.R.D. 159, 167 (1991). These administrative costs of fee shifting and their effects must also enter the analysis. See Bruce H. Kobayashi & Jeffrey S. Parker, No Armistice at 11: A Commentary on the Supreme Court's 1993 Amendment to Rule 11 of the Federal Rules of Civil Procedure, 3 Sup. Ct. Econ. Rev. 93 (1993); Polinsky and Rubinfeld, Sanctioning Frivolous Suits, *supra* note 3. These costs may militate not only in favor of simpler (but cruder) criteria for fee shifting but also in favor of using fee shifting sparingly. Courts might set fee-shifting thresholds so as to trigger fee shifting less frequently, in order both to reduce administrative costs and to increase the incentives of the parties to restrain their own expenditures on their attorneys' fees.

B. The Use of Sanctions Other than Fee Shifting

Under Rule 11, courts may impose sanctions other than fee shifting, such as fines or nonmonetary sanctions. As amended in 1993, Rule 11 explicitly authorizes "directives of a nonmonetary nature" and "an order to pay a penalty into court," although these sanctions were also available under Rule 11 prior to 1993.³¹ As we have noted, there is some debate over whether fee shifting is too severe a sanction and when courts should use other sanctions instead. Our analysis also has implications for this question.

Although the authors of the 1993 amendments apparently believed that fee shifting often exceeds what is necessary for adequate deterrence,³² the analysis in this article suggests that fee shifting need not ever deter too many suits. If the courts set the appropriate thresholds for fee shifting, then fee shifting will not lead to excessive deterrence in the sense of discouraging any meritorious suits.³³ In fact, the most important shortcoming of reliance on fee shifting alone is that it would be insufficient for effective deterrence of all frivolous suits. Our analysis identifies circumstances in which fee shifting under Rule 11 would prove to be inadequate as a sanction. In these cases, larger sanctions are warranted. Our analysis suggests two limitations on the ability of courts to create optimal incentives through the use of Rule 11 as a two-sided fee-shifting device.

One limitation flows from the inherent limits on the usefulness of fee-shifting rules in reducing excessive incentives to sue. If the amount at stake is sufficiently large relative to litigation costs, then condition (10) holds: given the probability of judicial error, the expected gain from bringing the marginal case exceeds the litigation costs of both sides. Thus, even the prospect of always paying the litigation costs of both sides would not discourage all frivolous suits. In these cases, a court would have to impose additional penalties in order to discourage these suits. The same analysis suggests how in theory courts might increase its sanctions just enough to reduce the expected value for the plaintiff of bringing the marginal suit down to zero.³⁴

³¹ See William W. Schwarzer, Sanctions under the New Federal Rule 11—a Closer Look, 104 F.R.D. 181, 201–4 (1985).

³² The Advisory Committee on Civil Rules conceded that "cost-shifting may be needed for effective deterrence" in some situations but sought "to emphasize that cost-shifting awards should be the exception, rather than the norm, for sanctions." Attachment B at 4, reprinted in 146 F.R.D. at 524.

³³ Fee shifting would still be more than sufficient to deter some frivolous suits, even if all meritorious suits have positive expected value for the plaintiff.

³⁴ Specifically, sufficiently large sanctions would replace the defendant's costs, C_d , in (10) with some greater penalty such that condition (10) fails to hold. Then this sanction, if

There is, however, another limitation on fee shifting under Rule 11. Courts generally reserve Rule 11 sanctions for parties that lose by wide margins, so that fees are not shifted in favor of a losing party. In terms of the model presented in this article, this practice restricts the thresholds, \bar{y}_p and \bar{y}_d , so that $\bar{y}_p < \bar{x} < \bar{y}_d$. We have seen, however, that given these constraints, fee shifting cannot ensure that plaintiffs have optimal incentives in all cases, even if litigation costs are large enough relative to the amount at stake to ensure condition (10) does not hold. Given this limitation, fee shifting under Rule 11 can provide optimal incentives if and only if the inequalities in (17) and (18) both hold strictly, that is, if and only if

$$C_p \theta_d < \theta_p D < C_p + C_d \theta_d.$$

That is, in the marginal case, the plaintiff's expected gain from adjudication must fall between its expected liability for litigation costs under the classic pro-plaintiff rule and its expected liability under the classic pro-defendant rule.³⁵ As we have noted, Rule 11 allows for sanctions other than the award of attorneys' fees, and these other sanctions may supplement the use of fee shifting. Courts may use these sanctions to address the problems that arise when either condition (17) or condition (18) fails to hold.

If litigation costs are too small relative to the amount at stake for inequality (18) to hold, then the prospect of Rule 11 fee shifting (subject to the constraint $\bar{y}_d \leq \bar{x}$) would be insufficient to deter all frivolous lawsuits, and additional penalties would be appropriate. That is, if the amount at stake is sufficiently large relative to litigation costs, so that the plaintiff will have excessive incentives to bring suit even under the classic pro-defendant rule, then the court would have to impose a larger sanction on a losing plaintiff than pro-defendant fee shifting. In such cases, one must supplement or replace C_d under such restricted fee shifting with some other sanctions for plaintiffs.³⁶ A sufficiently large sanction could reduce the expected value of the marginal suit to zero, so as to provide optimal incentives for the plaintiff, even if courts can impose sanctions only against losing plaintiffs.

If instead the plaintiff's litigation costs are so great relative to the

applied with the appropriate probability, would ensure that the plaintiff would have neither too much nor too little incentive to bring suit.

³⁵ Nevertheless, Rule 11 can provide optimal incentives to the plaintiff over a wider range of cases than either the American rule or the British rule, which subject fee shifting to still more severe restrictions.

³⁶ If (18) failed to hold, a court could substitute a greater sanction for C_d in (18) such that the inequality in (18) would hold.

amount at stake that condition (17) fails to hold, then too little litigation results even under the classic pro-plaintiff rule. The authors of Rule 11, however, did not focus on this problem and do not explicitly provide courts with any instrument to handle these cases. In these cases, our analysis suggests that a court should supplement pro-plaintiff fee shifting by awarding a winning plaintiff an extra amount.³⁷ Under Rule 11, a court might characterize such an award as a sanction imposed on the defendant for a frivolous defense against a meritorious claim by the plaintiff.

The authors of the most recent amendments to Rule 11, however, did not favor such sanctions. The committee notes explain:

Since the purpose of Rule 11 sanctions is to deter rather than to compensate, the rule provides that, if a monetary sanction is imposed, it should ordinarily be paid into court as a penalty. However, under unusual circumstances, . . . deterrence may be ineffective unless the sanction not only requires the person violating the rule to make a monetary payment, but also directs that some or all of this payment be made to those injured by the violation. Accordingly, the rule authorizes the court . . . to award attorney's fees to another party. Any such award to another party, however, should not exceed the expenses and attorneys' fees for the services directly and unavoidably caused by the violation. . . .³⁸

To the extent courts follow this approach, sanctions paid to the other party under Rule 11 would not go beyond fee shifting. Although our analysis would suggest that such sanctions may be useful in inducing plaintiffs to bring meritorious suits, the authors of Rule 11 were instead solely concerned with deterring frivolous papers.

VIII. CONCLUSION

We have shown that, when plaintiffs cannot predict the outcome of litigation with certainty, neither the American rule of litigation cost allocation (under which each litigant bears its own expenses) nor the British rule (under which the losing litigant pays the attorneys' fees of the winning litigant) would induce plaintiffs to make optimal decisions to bring suit. In particular, plaintiffs may bring frivolous suits when litigation costs are sufficiently small relative to the amount at stake, and plaintiffs may not bring some meritorious suits when litigation costs are sufficiently large relative to the amount at stake. We have analyzed the effects of more general fee-shifting rules that are based not only on the identity of

³⁷ Similarly, Polinsky and Rubinfeld suggest increasing awards to winning plaintiffs to encourage those with "legitimate" cases to sue. See Polinsky & Rubinfeld, *Sanctioning Frivolous Suits*, *supra* note 3; Polinsky & Rubinfeld, *Optimal Awards*, *supra* note 3.

³⁸ Fed. R. Civ. P. 11 committee notes, reprinted in 146 F.R.D. 583, 587-88 (1993).

the winning party but also on how strong the court perceives the case to be at the end of the trial—that is, the “margin of victory.” In particular, we have explored how and when one can design such a rule to induce plaintiffs to sue if and only if they believe their cases are sufficiently strong.

Our model shows how courts can use this additional instrument to improve the incentives for plaintiffs to bring suit. Our analysis also indicates the factors that courts would consider in deciding whether to shift fees if the objective is to induce optimal litigation decisions. We believe that this analysis suggests some considerations to guide courts applying sanctions under Federal Rule of Civil Procedure 11.

APPENDIX

PROOFS OF PROPOSITIONS

Proof of Proposition 1. Under the proposed rule, the plaintiff would sue if and only if

$$-C_p + \text{pr}(x_c > \bar{y}_p | x_p) C_p + \text{pr}(x_c > \bar{x} | x_p) D > 0.$$

Thus, the plaintiff would sue if and only if $x_p > s^*$, where s^* is defined by

$$-C_p + [1 - F(\bar{y}_p - s^*)] C_p + [1 - F(\bar{x} - s^*)] D = 0. \quad (\text{A1})$$

Thus, to ensure that $s^* = \bar{x}$, we must set \bar{y}_p such that

$$-C_p + [1 - F(\bar{y}_p - \bar{x})] C_p + \theta_p D = 0. \quad (\text{A2})$$

Note that the left-hand side of (A2) is monotonically decreasing in \bar{y}_p as long as $0 < F(\bar{y}_p - \bar{x}) < 1$. If we let \bar{y}_p approach ∞ , then $F(\bar{y}_p - \bar{x})$ goes to 1, so that the left-hand side of (A2) must be negative, and plaintiffs would not sue often enough. If we let \bar{y}_p approach $-\infty$, then $F(\bar{y}_p - \bar{x})$ goes to 0, so that the left-hand side of (A2) must become positive, so that plaintiffs would sue too often. Thus, there exists a \bar{y}_p that solves (A2). Specifically, solving (A2) for \bar{y}_p yields the expression for \bar{y}_p given in (7). Q.E.D.

Proof of Proposition 2. Under the proposed rule, the plaintiff would sue if and only if

$$-(C_p + C_d) + \text{pr}(x_c > \bar{y}_d | x_p) C_d + \text{pr}(x_c > \bar{x} | x_p) D > 0.$$

Thus, the plaintiff will sue if and only if $x_p > s^*$, where s^* is defined by

$$-(C_p + C_d) + [1 - F(\bar{y}_d - s^*)] C_d + [1 - F(\bar{x} - s^*)] D = 0. \quad (\text{A3})$$

Thus, to ensure that $s^* = \bar{x}$, we must set \bar{y}_d such that

$$-(C_p + C_d) + [1 - F(\bar{y}_d - \bar{x})] C_d + \theta_p D = 0. \quad (\text{A4})$$

Note that the left-hand side of (A4) is monotonically decreasing in \bar{y}_d as long as $0 < F(\bar{y}_d - \bar{x}) < 1$. If we let \bar{y}_d approach $-\infty$, then $F(\bar{y}_d - \bar{x})$ goes to 0, so that the left-hand side of (A4) must be positive, and plaintiffs would sue too often. If we let \bar{y}_d approach ∞ , then $F(\bar{y}_d - \bar{x})$ goes to 1. In this case, the left-hand side of (A4) becomes nonpositive if and only if the following inequality also holds:

$$\theta_p D \leq C_p + C_d. \quad (A5)$$

Thus, there exists a \bar{y}_d that solves (A4) if and only if (A5) holds; otherwise, there would always be too much litigation. Assuming such a \bar{y}_d exists, solving (A4) for \bar{y}_d yields the expression for \bar{y}_d given in (9). Q.E.D.

Proof of Proposition 3. Under the proposed rule, the plaintiff would sue if and only if

$$-C_p + \text{pr}(x_c > \bar{y}_p | x_p) C_p - \text{pr}(x_c \leq \bar{y}_d | x_p) C_d + \text{pr}(x_c > \bar{x} | x_p) D > 0.$$

Thus, the plaintiff will sue if and only if $x_p > s^*$, where s^* is defined by

$$-C_p + [1 - F(\bar{y}_p - s^*)] C_p - F(\bar{y}_d - s^*) C_d + [1 - F(\bar{x} - s^*)] D = 0. \quad (A6)$$

Thus, to ensure that $s^* = \bar{x}$, we must set \bar{y}_d and \bar{y}_p such that (16) holds. Note that the left-hand side of (16) is monotonically decreasing in \bar{y}_d as long as $0 < F(\bar{y}_d - \bar{x}) < 1$ and in \bar{y}_p as long as $0 < F(\bar{y}_p - \bar{x}) < 1$.

If we let both \bar{y}_d and \bar{y}_p approach ∞ , then both $F(\bar{y}_d - \bar{x})$ and $F(\bar{y}_p - \bar{x})$ go to 1, so that the left-hand side of (16) must become nonpositive, if and only if (A5) holds. If we let both \bar{y}_d and \bar{y}_p approach $-\infty$, then both $F(\bar{y}_d - \bar{x})$ and $F(\bar{y}_p - \bar{x})$ go to 0. In this case, the left-hand side of (16) must become positive, and plaintiffs would sue too often. Thus, there exists a pair (\bar{y}_d, \bar{y}_p) that solves (16) if (A5) holds. If (A5) does not hold, then there would always be too much litigation.

If such a solution exists, then there exists a whole family of solutions (\bar{y}_d, \bar{y}_p) . We can solve (16) for either \bar{y}_d or \bar{y}_p . The solution for \bar{y}_d may be expressed as a monotonically decreasing function of \bar{y}_p :

$$\bar{y}_d(\bar{y}_p) = \bar{x} + F^{-1}\{[\theta_p D - F(\bar{y}_p - \bar{x}) C_p] / C_d\}. \quad (A7)$$

Similarly, the solution for \bar{y}_p may be expressed as a monotonically decreasing function of \bar{y}_d :

$$\bar{y}_p(\bar{y}_d) = \bar{x} + F^{-1}\{[\theta_p D - F(\bar{y}_d - \bar{x}) C_d] / C_p\}. \quad (A8)$$

Thus, as long as both $0 < F(\bar{y}_d - \bar{x}) < 1$ and $0 < F(\bar{y}_p - \bar{x}) < 1$, increases in either \bar{y}_d or \bar{y}_p would substitute for increases in the other. Q.E.D.

